

GPLSI Emotion Analysis v1.0: Análisis de emociones en textos

GPLSI Emotion Analysis v1.0: Emotion analysis on texts

Resumen de la aplicación: GPLSI Emotion Analysis v1.0 constituye una interfaz de programación de aplicaciones (API) que incluye librerías de programación útiles para sistemas de terceros. El objetivo de esta tecnología es analizar textos para detectar y clasificar las emociones expresadas. La Detección y Clasificación de Emociones en textos permite identificar y clasificar las emociones expresadas en un texto. La herramienta aquí descrita es capaz de discriminar entre el siguiente conjunto de emociones: alegría, tristeza, disgusto, miedo y sorpresa. Esta funcionalidad implica dos niveles de granularidad: emoción a nivel global (ENG) y emoción a nivel de aspectos (ENA). La ENG del texto se define como el análisis de todo el texto para determinar el estado de emoción general. La ENA determina el estado de emoción para un aspecto concreto de una entidad dentro un contexto determinado. Por ejemplo, en un mismo texto (usualmente comentarios de usuarios) se pueden expresar emociones diferentes y opuestas hacia los actores de una película y su banda sonora. Dichas funcionalidades, ENG y ENA, se basan en un algoritmo SVM (Máquina de Soporte Vectorial). Esta tecnología aplica skip-grams para extraer patrones de los textos y así compararlos con otros casos (conjunto de entrenamiento) en los que de manera segura se expresan emociones. El conjunto de entrenamiento empleado en esta herramienta ha sido: Potter, Grimms and HCAndersen. Los resultados del estado del arte se posicionan alrededor del 52% de F1. En el caso del sistema descrito aquí, se obtiene de precisión entre 54% y 65% de F1 bajo las mismas condiciones. Es importante resaltar que los clasificadores SVM se pueden actualizar y/o reemplazar para su uso en otros dominios al re-entrenarlos con otros conjuntos de entrenamientos.

Summary: Summary: GPLSI Emotion Analysis v1.0 is an application programming interface (API) which involves programming libraries for third-parties. The aim of this technology is to analyse texts to detect and classify emotions. The emotions detected by the technology described here are: happiness, sadness, disgust, fear and surprise. This task involves two types of analysis: global emotion analysis (GEA) and aspect-based emotion analysis (ABEA). GEA allows detecting and inferring the sentiment emotion expressed in the whole text, whereas ABEA determines the sentiment emotion regarding particular aspects mentioned in it. Both functionalities are based on a Support Vector Machines (SVM) algorithm. This technology applies skip-grams techniques in order to extract lexical patterns to feed a SVM algorithm and then learn to identify and classify emotions in text. The training dataset used has been obtained from Potter, Grimms and HCAndersen. The state-of-the-art performance in Emotion Detection is set around 52% of F1. The technology described here obtains a precision between 54% and 65% of F1 under the same conditions. It is important to highlight that the SVM algorithm can be updated and/or replaced if necessary to work in other domains.

Especificaciones técnicas/ Technical Specifications

Lenguaje de programación/ Development Language: Java

Entorno Operativo/ Operating Environment: Linux / Windows

Versión/Version: 1.0

Estructura de ficheros/Folder structure:

Directorio/Folder ->

sentiment-analysis-core:

- nbactions-giventext.xml
- nbactions.xml
- pom.xml
- nbactions-multitextscommand.xml
- nb-configuration.xml

Directorio/ Folder ->

sentiment-analysis-core/src/main/resources/ models:

- en-chunker.bin
- en-ner-date.bin
- en-ner-money.bin
- en-ner-percentage.bin
- en-ner-time.bin
- en-pos-maxent.bin
- en-sent.bin
- testmallet.txt
- en-ner-asset.bin
- en-ner-location.bin
- en-ner-organization.bin
- en-ner-person.bin
- en-parser-chunking.bin
- en-pos-perceptron.bin
- en-token.bin

Directorio/ Folder ->

sentiment-analysis-core/src/main/resources/t raining:

- emotion.data

Directorio / Folder ->

sentiment-analysis-core/src/main/java/es/ua/ sentimentanalysis:

- Main.java

Directorio / Folder ->

sentiment-analysis-web:

- nb-configuration.xml
- pom.xml

Directorio / Folder ->

sentiment-analysis-web/src/main/java/es/ua/

- Subdirectorio / Subfolder -> analyzer

- Definition.java
- MorphoSyntacticAnalysis.java
- POSTagger.java
- TextEntity.java
- SentimentAnalyzer.java

- Subdirectorio / Subfolder ->

contextsimilarity

- ContextSimilarityFactory.java
- ContextSimilarityStrategy.java

- Subdirectorio / Subfolder ->

sentiment

- EmotionAttributes.java
- EmotionDetector.java
- EmotionReader.java
- SentimentAnalysis.java
- SentimentAttributesEmotionS
cores.java
- SentimentAttributes.java
- SentimentAttributesSocial.jav
a
- SentimentAttributesSocialSco
re.java
- SentimentSocialAttributesEm
otionScoresResponse.java
- SentimentSocialAttributesRes
ponse.java

- Subdirectorio / Subfolder -> util:

- GsonInstance.java
- LookUpEntity.java
- PorterStemmer.java
- LevenshteinCalculator.java
- Pair.java

sentimentanalysis/application:

- ApplicationConfig.java

Directorio / Folder ->

sentiment-analysis-web/src/main/java/es/ua/ sentimentanalysis/services:

- SentimentAnalysis.java
- Subdirectorio / Subfolder -> exception

- InvalidParametersException.java
- **Subdirectorio / Subfolder -> filter**
 - EntityFinderFilter.java
- **Subdirectorio / Subfolder -> internal/annotation**
 - ApplicationInfo.java
- **Subdirectorio / Subfolder -> internal/utills**
 - AnnotatedTextProcessor.java
 - ReflectionUtils.java
- **Subdirectorio / Subfolder -> listener**
 - ContextListener.java
- **Subdirectorio / Subfolder -> parameters**
 - IParameters.java
 - AnaysisSentimentParameters.java
- **Subdirectorio / Subfolder -> security**
 - SecurityUtils.java
 - UntrustedAccessExceptionMapper.java
 - UntrustedAccessException.java

Directorio / Folder ->

sentiment-analysis-web/src/main/resources/models:

- en-chunker.bin
- en-ner-date.bin
- en-ner-money.bin
- en-ner-percentage.bin
- en-ner-time.bin
- en-pos-maxent.bin
- en-sent.bin
- testmallet.txt
- en-ner-asset.bin
- en-ner-location.bin
- en-ner-organization.bin
- en-ner-person.bin
- en-parser-chunking.bin
- en-pos-perceptron.bin
- en-token.bin

Directorio / Folder ->

sentiment-analysis-web/src/main/resources/training:

- emotion.data

Directorio / Folder ->

sentiment-analysis-web/src/main/webapp:

- entityfinder.html
- index.html
- logs.html
- **Subdirectorio / Subfolder -> css**
 - bootstrap.min.css
 - bootstrap-theme.min.css
 - log-style.css
 - main.css
 - style.css
- **Subdirectorio / Subfolder -> doc**

- **Subdirectorio css**
 - reset.css
 - screen.css
- **Subdirectorio / Subfolder -> img**
 - bg.png
 - logo.png
 - pet_store_api.png
 - wordnik_api.png
 - explorer_icons.png
 - logo_small.png
 - throbber.gif
- index.html.bak
- o2c.html
- swagger-ui.min.js
- favicon.ico
- index.html
- **Subdirectorio / Subfolder -> lib**
 - backbone-min.js
 - jquery.slideto.min.js
 - swagger.js
 - handlebars-1.0.0.js
 - jquery.wiggle.min.js
 - swagger-oauth.js
 - handlebars-2.0.0.js
 - marked.js
 - underscore-min.js
 - highlight.7.3.pack.js
 - **Subdirectorio shred**
 - content.js
 - underscore-min.map
 - jquery-1.8.0.min.js
 - shred.bundle.js
 - jquery.ba-bbq.min.js
 - swagger-client.js
 - swagger-ui.js
- **Subdirectorio / Subfolder -> fonts**
 - glyphsicons-halflings-regular.eot
 - glyphsicons-halflings-regular.ttf
 - glyphsicons-halflings-regular.svg
 - glyphsicons-halflings-regular.woff
- **Subdirectorio / Subfolder -> img**
 - bg-header.png
 - bg.png
 - gplsi.png
- **Subdirectorio / Subfolder -> js**
 - ajax.js
 - bootstrap.min.js
 - jquery-2.1.1.min.js
 - logs.js
 - app.js
 - **Subdirectorio controllers**
 - main.js
 - jquery-ui.js

- **Subdirectorio / Subfolder -> META-INF**
 - context.xml
- **Subdirectorio / Subfolder -> views**
 - main.html
 - **Subdirectorio / Subfolder -> fragments**
 - footer.html
- header.html
- **Subdirectorio / Subfolder -> WEB-INF**
 - host.properties
 - log4j.properties
 - security.properties
 - test.txt
 - web.xml

Requerimientos: Java 1.7 o superior instalado, 3 GB de RAM o superior, 1 GB de disco duro libre en sistema, Apache Tomcat 8.0 o superior y Maven. Además, son necesarios e imprescindibles los modelos para realizar el análisis de emociones (i.e.

“**emotion.data**”), así como los modelos de Apache OpenNLP¹. Los modelos OpenNLP ya se incluyen en la distribución de la aplicación, por comodidad del usuario, pero no necesariamente se tiene que distribuir dentro de ésta, ya que son recursos externos.

Instalación: Descargar el código fuente del repositorio de control de versiones. Se deben generar los binarios tanto de la aplicación de consola como web utilizando Maven. Antes de generar los binarios es importante asegurarse de que los modelos de análisis de emociones así como los de Apache OpenNLP se encuentran en los directorios [core|web]/src/main/resources/training y [core|web]/src/main/resources/models, respectivamente.

En ese momento la aplicación de consola está lista para funcionar.

La instalación de la aplicación web es opcional y, además de lo anterior, es necesario desplegar el war generado en el servidor de aplicaciones Tomcat.

En principio se puede usar cualquier sistema operativo, pero sólo ha sido probado en Ubuntu².

Ejecución: La aplicación web dispone de una página de documentación del servicio web REST. El servicio se puede ejecutar directamente desde esta documentación en un navegador o mediante el uso clientes para el servicio web REST.

La aplicación de consola dispone de un comando para detectar las emociones expresadas en textos en idioma Inglés, así como un comando de ayuda. A continuación mostramos algunos ejemplos:

- **[\$[nombre del jar] help analyze**
 - Muestra la ayuda del comando, con los parámetros y su descripción.
- **[\$[nombre del jar] analyze -f/--file [nombre del archivo txt]**
 - Permite proporcionar un archivo con el texto que se desea analizar. Por defecto, el análisis de polaridad se realiza solo a nivel global. Cada texto debe ser un párrafo dentro del archivo, separado por un salto de línea. Para más

¹ Disponibles en <http://opennlp.sourceforge.net/models-1.5/>

² <https://www.ubuntu.com/>

información ver la ayuda mencionada en el ejemplo anterior.

Dependencias de programas de terceros:

El listado completo de dependencias, tanto para la aplicación de consola como para el servicio web, se encuentra en los archivos pom.xml. Ahí se incluye un listado de las dependencias que se descargarán de varios repositorios Maven, en el cual las librerías propias del GPLSI se marcan con un * y en negrita:

sentiment-analysis-core 1.0.0

- airline 0.7
- commons-lang 3-3.1
- gson 2.3
- guava 19.0
- jettison 1.3.3
- junit 4.12
- lucene-analyzers-common 4.9.0
- lucene-queryparser 4.9.0
- opennlp tools 1.5.3
- **sam-sentiment-core 1.1.0***
- annotations 2.0.3
- **classifier-base 1.0.0***
- commons-io 2.4
- hamcrest-core 1.3
- jackson-annotations 2.5.0
- jackson-core 2.5.0
- jackson-databind 2.5.0
- java-cup 0.11a
- javax.inject 1
- jwnl 1.3.3
- lucene-core 4.9.0
- lucene-queries 4.9.0
- lucene-sandbox 4.9.0
- opennlp maxent 3.0.3
- slf4j-api 1.7.21
- slf4j-api 1.0.1
- stax-api 1.0.1
- weka-stable 3.6.14

Web

- **sentiment-analysis-core 1.0.0***
- gson 2.3
- guava 15.0
- jackson-annotations 2.5.0
- jackson-core 2.5.0
- jackson-databind 2.5.0
- jackson-jaxrs-base 2.5.0
- jackson-jaxrs-json-provider 2.5.0
- jackson-jaxrs-xml-provider 2.5.0
- jackson-module-jaxb-annotations 2.5.0
- javaxee-web-api 7.0
- javassist 3.18.1-GA
- javax.annotation-api 1.2
- javax.servlet-api 3.0.1
- javax.ws.rs-api 2.0.1
- jaxb-api 2.2.7
- jersey-client 2.15
- jersey-common 2.15
- jersey-container-servlet 2.15
- jersey-container-servlet-core 2.15
- jersey-entity-filtering 2.15
- jersey-guava 2.15
- jersey-media-json-jackson 2.15
- jersey-media-multipart 2.15
- jersey-server 2.15
- log4j 1.2.17
- persistence-api 1.0
- swagger-annotations 1.3.11
- swagger-core_2.10-1.3.11
- swagger-core 1.5.3
- swagger-jaxrs_2.10-1.3.11
- swagger-jersey2-jaxrs_2.10-1.3.11
- swagger-utils_2.10-1.3.11
- swagger-models 1.5.3
- validation-api 1.1.0
- xom 1.2.10
- airline 0.7
- annotations 2.0.3
- aopalliance-repackaged 2.4.0-b06
- **classifier-base 1.0.0***
- commons-io 2.5
- commons-lang3-3.1
- commons-lang 2.4
- dom4j 1.6.1
- hamcrest-core 1.3
- hk2-api 2.4.0-b06
- hk2-locator 2.4.0-b06
- hk2.hk2-utils 2.4.0-b06
- jackson-dataformat-xml 2.5.0
- jackson-module-jsonSchema 2.1.0
- jackson-module-scala 2.10-2.41
- java-cup 0.11a
- javax.inject 2.4.0-b06

- javax.inject 1
- joda-convert 1.2
- joda-time 2.2
- json4s-ast_2.10 3.2.9
- json4s.json4s-core_2.10 3.2.9
- json4s.json4s-ext_2.10 3.2.9
- json4s.json4s-jackson_2.10 3.2.9
- json4s.json4s-native_2.10 3.2.9
- jsr305 2.0.1
- jsr311-api 1.1.1
- junit 4.12
- jwnl 1.3.3
- lucene-core 4.9.0
- lucene-queries 4.9.0
- lucene-queryparser 4.9.0
- lucene-analyzers-common 4.9.0
- lucene-sandbox 4.9.0
- mimepull 1.9.3
- opennlp-maxent 3.0.3
- opennlp-tools 1.5.3
- osgi-resource-locator 1.0.1
- paranamer 2.6
- reflections 0.9.9-RC1
- **sam-sentiment-core 1.1.0***
- scala-compiler 2.10.0
- scala-library 2.10.0
- scala-reflect 2.10.4
- scalap 2.10.0
- slf4j-api 1.6.3
- stax-api 1.0.2
- stax2-api 3.1.1
- swagger-annotations 1.5.3
- woodstox-core-asl-4.4.0
- xalan 2.7.0
- xercesImpl 2.8.0
- xml-apis 1.3.03
- weka-stable-3.6.14

Referencias/References

- [GTF15] Gutiérrez, Y.; Tomás, D.; Fernández, J. Benefits of Using Ranking Skip-Gram Techniques for Opinion Mining Approaches. Proceedings of eChallenges 2015 e-2015. pp 1-10. 2015
- [OP13] G. Özbal, D. Pighin, Evaluating the impact of syntax and semantics on emotion recognition from text. In International Conference on Intelligent Text Processing and Computational Linguistics (pp. 161-173). Springer Berlin Heidelberg. 2013